United States Department of Agriculture

NRCS Natural Resources
Conservation Service

Soil Data Delivery

And

Distribution

Outline Physical Design

December 20, 2001

Executive Summary

This document is the second stage in the development cycle for a comprehensive soil data delivery and distribution system. The first document, the Draft Requirements Statement (DRS), approved October 30, 2001, described the Soil Survey Division's requirements for a data delivery system. Now, the Outline Physical Design (OPD) proposes a strategy for meeting those requirements by identifying the major components of such a system and an approximate timeline for their development.

The development schedule for this system projects a warehouse database operational in September, 2002, so that State Soil Scientists can begin moving their approved data into it and delivering SSURGO products through the new system. By March, 2003, reporting and interpretation capabilities will be available, followed by other tools as resources permit.

It is possible to meet this aggressive schedule because several key parts of the data delivery system have already been developed or prototyped. The design calls for making maximum use of existing systems, especially NASIS and SSURGO. The major capabilities to be provided in the new system include:

- A Soil Data Warehouse serving as a single source for official soil data.

- A Data Mart to supply data for the Field Office Technical Guide, Section 2, including SSURGO datasets.

- A Data Mart to provide interactive interpretation and reporting capabilities from the official data.

- A Data Mart for exporting data to a variety of users.

- An Application Programming Interface to allow models and other application programs to access the official data in a secure manner.

The schedule requires that funding for hardware/software purchases and development staff be available in a timely manner. Cost estimates are very rough in this early stage of the analysis. This table summarizes the estimated costs by fiscal year.

Item	FY 2002	FY 2003	After FY 2003
Hardware/Software	$300,000 - 450,000	$150,000 – 250,000	
Development Staffing	$300,000 – 500,000	$200,000 – 300,000	
Operation and Maintenance		$400,000 – 500,000	$400,000 – 500,000
Total	$600,000 – 950,000	$750,000 – 1,050,000	$400,000 – 500,000

Contents of the Outline Physical Design

This document provides a overview of the proposed design of a system to meet the requirements described in the Draft Requirements Statement (DRS). At this stage in the design lifecycle a set of major components, or design units, is identified and described. Each system requirement is associated with the design unit or units that will fulfill it, so that the design can be evaluated. The major sections of the document are:

Overview Data Flow Diagram

Design Unit Descriptions

Data Model Issues

Review of Requirements

Design Unit Sequence

Conversion Approach

Summary of Conclusions

Interactions with Other Systems

A key insight guiding this OPD is that the data delivery system can be brought on line much more quickly by leveraging existing system components than by starting fresh with a complete new system design. Some of the existing systems are:

NASIS 5.0: Since the NASIS central server was brought on line in April, 2001, a major source of data for a warehouse has become easily accessible and NRCS has gained experience with managing a secure database of this size. In addition, reporting and interpretation capabilities in NASIS are similar to those called for in the DRS. Much of this can readily be adapted if a NASIS like architecture is used for the warehouse.

SSURGO: In 2001 a new format for SSURGO was adopted which is compatible with the Customer Service Toolkit. By providing this product through a FOTG Data Mart a number of high priority needs can be met.

Web Soil Data Viewer: A prototype of a Web based tool for geospatial analysis of soil data has been demonstrated. By linking this tool to a soil data warehouse a valuable new capability can be presented to the public with little new development cost.

Natural Resources Data Gateway: Another recently developed prototype, the Gateway provides a convenient way to locate and access soil data sets, and can be linked to the FOTG Data Mart.

MUIR: The Mapunit Interpretation Record continues to be a source of data for analysis and delivery to a number of customers even though the data are out of date. These capabilities can be converted to use the warehouse and provide a transition to current approved data. If there is a business need for it, the outdated soils data now in MUIR could also be retained in the warehouse as an old version.

Overview Data Flow Diagram

Design units are shown enclosed in dashed lines.
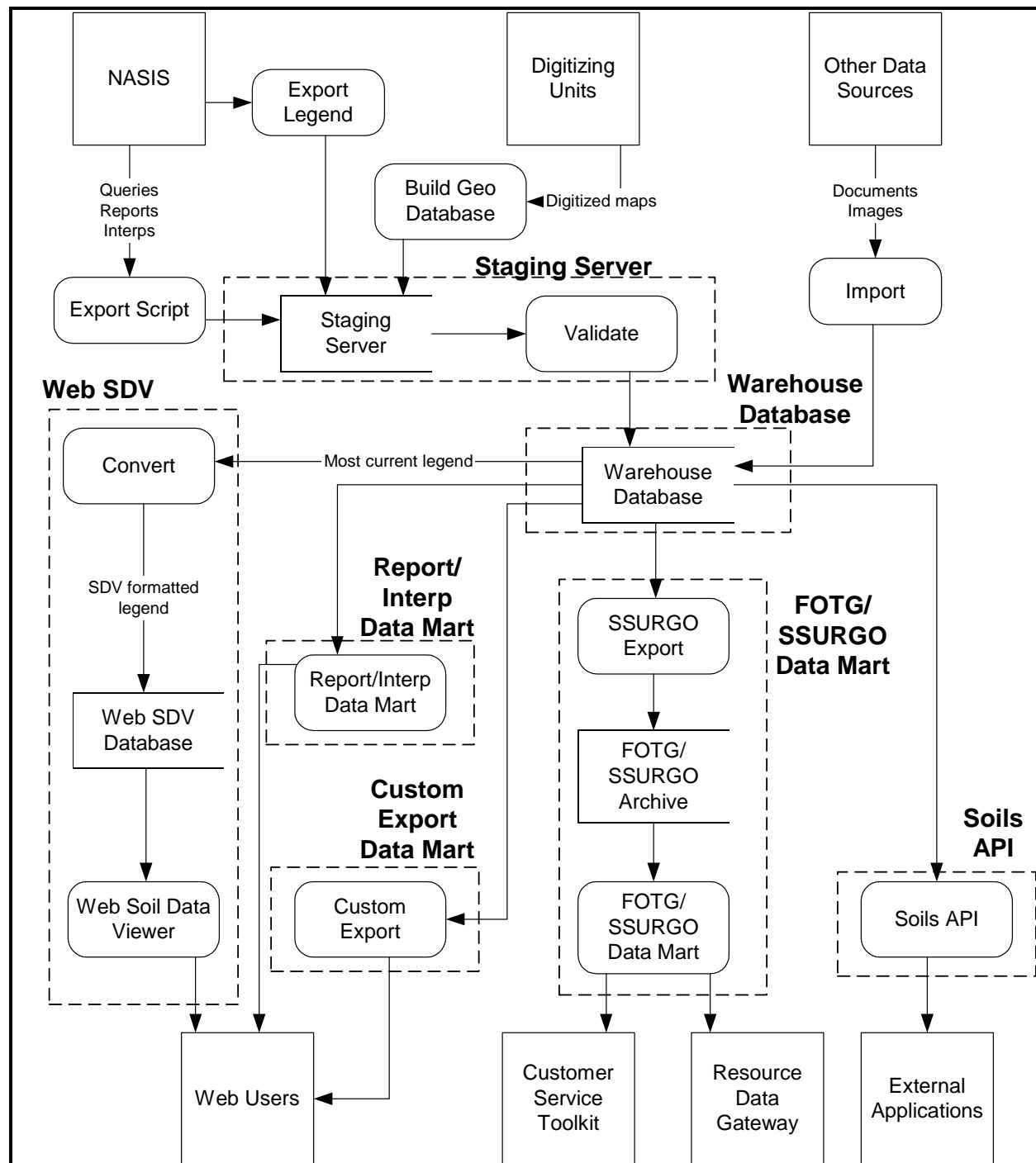
Design Unit Descriptions

The description of each design unit is in a summary form. Capacity estimates are preliminary, based on experience with existing systems such as SSURGO and MUIR. The next stage of analysis, the Total Requirements Statement, will elaborate the detailed, step by step processing to be performed and the detailed data models for the databases.

Warehouse Database

Contents:

- Soil data tables including all NASIS data elements, for all traditional (non-MLRA) soil surveys that have been approved for public release by State Soil Scientists.

- Digital soil maps, where available, for the same soil surveys.

- Other types of soil surveys, such as STATSGO or MLRA legends.

- Stored interpretation ratings generated from data and criteria in effect at the time a soil survey is approved for release.

- Prior versions of soils tables and maps, when superseded by a newer copy. Each version is identified by an effective date.

- Metadata including the contents of the NASIS data dictionary, spatial metadata, and other information needed to meet metadata requirements such as FGDC.

- Interpretation criteria, reports and queries similar to those used in NASIS.

- Supporting data including pedon descriptions, lab characterization data, performance measurements, field notes or images, as approved for public distribution by State Soil Scientists.

- Digitized soil survey area boundaries.

- Documents such as soil survey manuscripts, official series descriptions, soil taxonomy, FOTG tables, handbooks and manuals, when available in digital form.

Capabilities:

- Maintain linkage between soil map features and tabular data.

- Support query and retrieval of data by either geographic or tabular attributes.

- Identify locations where tabular data or digital maps are not available to complete a national coverage, or where multiple maps exist for one area.

- Prevent modification of data other than by addition of a new version.

- Provide backup, mirroring, load balancing and other security measures to protect the integrity of the data.

- Permit conversion of the data structure from time to time to meet business requirements.

Capacity:

- A national set of tabular and spatial soil survey data will require about 100 GB of disk space. Adding space for other data, plus software, redundant storage, and work space yields a estimate of 300 to 500 GB for an initial configuration.

- The warehouse database is visualized as a data source for the Data Marts, rather than supporting direct access by users. This configuration is similar to the NASIS database server, so a system of similar performance capacity is projected.

- The warehouse database will be available around the clock, but downtime for scheduled maintenance of a few hours per week at non-peak hours is acceptable. Disaster recovery time of 1 to 2 weeks is acceptable.

Options:

- We recommend storing the NASIS calculated interpretation results, as well as the interpretation criteria, when a survey is placed in the warehouse. In principle the same interpretation results could be generated from the warehouse on demand, by using the data and criteria that were in effect at the time of survey certification. However, there are complications in ensuring that the correct versions of the interpretation criteria are in the warehouse when a survey version is certified. Since the intent is to produce the same results for a survey each time they are requested (until a new version is certified) and the amount of storage is not excessive, it makes sense to store the results.

- The Spatial Database Engine (SDE) from ESRI, Inc. is the recommended product for storing the digital maps, since it can support a complete national coverage and linkage to the tabular data. No other products are comparable.

- The recommended database under SDE is Informix, to facilitate the use of custom interpretation and reporting software originally written for NASIS. Use of other database software, such as Oracle or SQL Server, would require porting of the NASIS tools or development of new tools, and result in a delay of up to a year in implementing these core capabilities. The potential savings in initial software costs for a solution such as SQL Server are more than offset by the increased development costs.

  The recent acquisition of Informix, Inc. by IBM casts doubts on the future of the Informix product. Because of the large Informix customer base, we are assuming that IBM will either continue to support Informix or provide a reasonable migration path. The NASIS database would also be affected by such a migration, and using the same architecture for both systems would reduce the overall cost of maintenance.

Staging Server

Contents:

- Transient storage of soil survey tabular data and maps in a form similar to that in the warehouse database.

- Transient storage of query, report and interpretation criteria for testing.

- No permanent storage of data.

Capabilities:

- Accept certified soil survey data from NASIS and digitized maps from a Digitizing Unit.

- Assign linkage (mapunit key) to maps and verify referential integrity with tabular data.

- Query, reporting and map display for data quality verification.

- Transfer data to warehouse database.

- Edit and test reports and interpretations (see Soil Reports and Interpretations Data Mart)

- Not open to the public. Access limited to authorized personnel.

Capacity:

- Storage of not more than 50 soil surveys at one time, plus associated data, requires 10 to 20 GB of disk space.

- The system should support 20 to 30 simultaneous users performing complex database and spatial analysis.

- Minimal usage is expected outside of normal working hours in North America.

Options:

- This represents a change in operating procedures for the Digitizing Units. Depending on the time required to design and implement the new procedures, the staging server might not be included in the first phase of warehouse development. Instead, certified surveys could be moved directly from NASIS to the warehouse, then exported for SSURGO certification in the same way they are now. The corresponding spatial data could be loaded into the warehouse from the SDE database currently maintained at NCGC. That would include surface point and line features not documented in NASIS.

FOTG / SSURGO Data Mart

Contents:

- Soil data tables, and associated digital maps where available, in SSURGO format.

- Current certified data for each soil survey as well as previously released versions as available.

- List of customers who receive data and wish to be notified of updates.

- Index of available FOTG data and documents.

Capabilities:

- Linkage to Natural Resources Gateway for locating available data sets.

- Web based access.

- Extract data from warehouse database, converting tabular and map data to SSURGO format.

- Download data for use in the Soil Data Viewer and other components of the Customer Service Toolkit.

- Download data to customers other than Service Centers, like the existing SSURGO distribution.

- Download tabular data without maps when digital mapping is not available.

- Provide access to other documents from the warehouse as appropriate for the Field Office Technical Guide, Section 2, including digital manuscripts, images, etc.

- Send notices of data updates to registered customers.

Capacity:

- Archived data sets for the nation will require about 100 GB of disk space initially.

- There are currently about 200 SSURGO downloads from Fort Worth per day, which gives a lower bound for the capacity of this data mart.

- The mart will be available around the clock, with scheduled downtime acceptable.

Options:

- In principle a SSURGO package for a survey area could be extracted from the warehouse database on demand, but it is likely to be a resource intensive process. This mart is

designed to hold current and past SSURGO packages, as well as SSURGO formatted data tables where maps are not available, in an archived form for ready distribution.

- Some aspects of the FOTG need further analysis. For example, it is likely that data and documents in the warehouse will need to be available for online interaction in addition to downloading. The Web SDV Data Mart may be able to satisfy some of these needs. Another possibility to be investigated is automatic delivery (push) of updated FOTG data to the appropriate service centers rather than downloading on request.

- The FGDC standard for soils data needs to be updated, and it is hoped that it will be close to SSURGO in content. If so this data mart should be capable of delivering an export in FGDC format.

Web Soil Data Viewer

Contents:

- Current versions of soil surveys that have certified data and digital maps.

- A geospatial database separate from the warehouse database, using the architecture that supports the Web SDV, including SQL Server, ArcIMS, and Microsoft IIS.

Capabilities:

- Extract latest version of soil surveys from warehouse database and convert to appropriate format for Web SDV.

- Host the Web SDV application.

Capacity:

- The USDA Lighthouse project recently collected performance data that can be used to estimate processor requirements for this data mart. The number of simultaneous users is hard to estimate, and can be expected to grow over time as the service becomes known. Scalability of the system will be a major design goal.

- A complete national database for this application will require about 100 GB of disk space.

- The mart will be available around the clock, with scheduled downtime acceptable.

Options:

- Eventually the Web SDV program could be converted to use the API described below. This data mart is intended to take immediate advantage of the prototype Web SDV.

- The geodatabase in this data mart could be made available for online geographic analysis to users having appropriate client software, such as ArcView.

Soil Reports and Interpretations Data Mart

Contents:

- Web based interface for selecting an area of interest (by legend, mapunit, or arbitrary boundary) and type of output.

- Prewritten report scripts and interpretation criteria.

Capabilities:

- Generate interpretations and reports using query, report and interpretation scripts stored in the warehouse database. This includes applying new interpretation criteria to existing data.

- Deliver report output as HTML page, Postscript print file, or plain text file.

- Interface with GIS for spatial querying or map generation.

- Access any version of data in the warehouse database with appropriate query.

- Report on changes between versions.

Capacity:

- Processor load on this data mart is likely to be similar to that on the NASIS application servers. The number of simultaneous users will grow over time, so scalability of the configuration will be an important requirement.

- Disk storage is needed only for software and working space, about 10 GB per server.

- The mart will be available around the clock, with scheduled downtime acceptable.

Options:

- This is intended to make use of existing NASIS tools for report and interpretation generation, to provide quick implementation. This does not preclude development or purchase of other tools at a later time.

- Because of the requirement to maintain past versions of data in the warehouse database, its structure will differ from that of NASIS, particularly in the key columns. Although some kinds of key changes can be supported transparently by NASIS, there may be some reports or interpretations that cannot run unchanged on both NASIS and the warehouse. Since modification of data, including report and interpretation scripts, is not permitted in the warehouse, a different environment is needed for testing and modifying these scripts. It is likely that the Staging Server can support this need.

- A set of reports to identify changes between versions of data in the warehouse is required. This cannot be addressed with NASIS style reports, so a special report program would have to be developed.

Custom Export Data Mart

Contents:

- Data dictionary driven web interface for custom exports.

- User saved export configurations and standard export configurations.

Capabilities:

- Identify a set of map units and a set of attributes to produce a table or tables of data to be downloaded.

- Provide common functions for collapsing data into simpler structures, such as dominant component, surface layer, first restriction, representative value, top level interpretation, etc.

- Selection of output formats to match input requirements of various tools or models.

- Selection of metadata content and format.

- Delivery of supporting data (pedons, lab data, documents, etc.).

Capacity:

- The processing and storage requirements for this mart are similar to those for reporting and interpretations. It should be possible to share servers, with the option of adding servers and load balancing as the traffic requires.

- There are currently about 40 downloads per day from the ISU site, primarily MUIR and OSD files.

Options:

- This is intended to replace and extend the capabilities of the current MUIR download. An initial implementation could be to just reproduce the output options provided in the current MUIR site.

- This is an opportunity for incremental development, beginning with a small set of standard export options, and gradually adding more capabilities.

Soils Application Programming Interface (API)

Contents:

- A documented set of software interfaces for retrieving data from the warehouse database for use in models or third party software packages.

Capabilities:

- Allow application programs to access the soils data online while preserving all security requirements and business rules.

- Insulate applications from some of the internal complexity of the warehouse database. For example, the API could provide access only to the current version of any soil survey, or present a view that matches an older database structure for a period of time after a structure change takes place.

- Allow different views for authorized users and the general public.

Capacity:

- The processing and storage requirements for this mart are similar to those for reporting and interpretations. It should be possible to share servers, with the option of adding servers as the traffic requires.

Options:

- One approach would provide ODBC access to a database consisting of views into the warehouse database. This is relatively simple to set up, but requires the application programmers to do most of the work of navigating through the soils tables. Some business rules could not be enforced with this approach.

- A more extensive API could provide a number of capabilities to make the database easier to use, such as aggregating repeating group data, selecting representative values, identifying dominant components, and so forth. Enterprise Java Beans are a possible implementation technique.

Data Model Issues

Development of a detailed data model will begin in the next stage of analysis. The generalized data model from the DRS is reproduced below for reference. At this stage several design issues can be identified which will affect the data model, some of which may require business decisions.

- Tabular data now stored in NASIS, such as legends, data mapunits and pedons, will maintain nearly the same data structure in the warehouse database. Structure differences may be needed where data in NASIS are not appropriate for the warehouse, such as program management data. In addition, since the warehouse can maintain more than one version of a NASIS record, version identifiers and a new system of primary keys will need to be developed.

- At this time it appears that the basic unit for versioning will be what is called in NASIS an "object". Objects include legends, data mapunits, sites, pedons, reports, rules, etc. It is not practical to place versions of individual tables or rows in the warehouse and still be able to quickly reconstruct a whole object version. On the other hand, checking a full object version into the warehouse does not provide any information on exactly what has changed since the last version, and requires more space. Our preliminary conclusion is that a request to extract a complete object version is more time critical than a request to identify detailed changes between versions.

- In a geographic database as proposed for the warehouse, soil map polygons and other features are stored in relational database tables. These tables need to be designed and added to the data model, making use of the prototype geodatabase developed at NCGC. The geographic features to be included in the model (and their related tabular data) are:

    o State, county, survey area boundary polygons (Area table)

    o Soil delineation polygons and lines (Mapunit table)

    o Site points (Site table)

    o Spot symbol points and lines (not in NASIS)

    o PSU points and polygons (NRI)

- The data model for the soil characterization (lab) data has not been designed. It also remains to be decided if there is a need to put lab data into NASIS as well as the warehouse. The data will not be edited after exporting from the lab (LIMS) database, so it technically should not be in the transactional NASIS database. But there may be justification for keeping a copy in NASIS to simplify analysis and reporting for NASIS users.

- Soil classification is currently stored in a relational database independent of NASIS. A business goal is to develop closer integration of the classification and the official series with the transactional NASIS database, and to include them in the warehouse database.

Further analysis is needed in this area. A short term solution could be to incorporate the existing SC and OSD web sites into one of the Data Marts.

- The requirements call for storage of documents, images and other files not now in NASIS. Detailed design work is needed for this also. In many cases these items can be associated with a NASIS object, such as a legend, and can be accommodated in the database by adding new tables for documents and images.

Soil Data Delivery and Distribution
Generalized Data Model
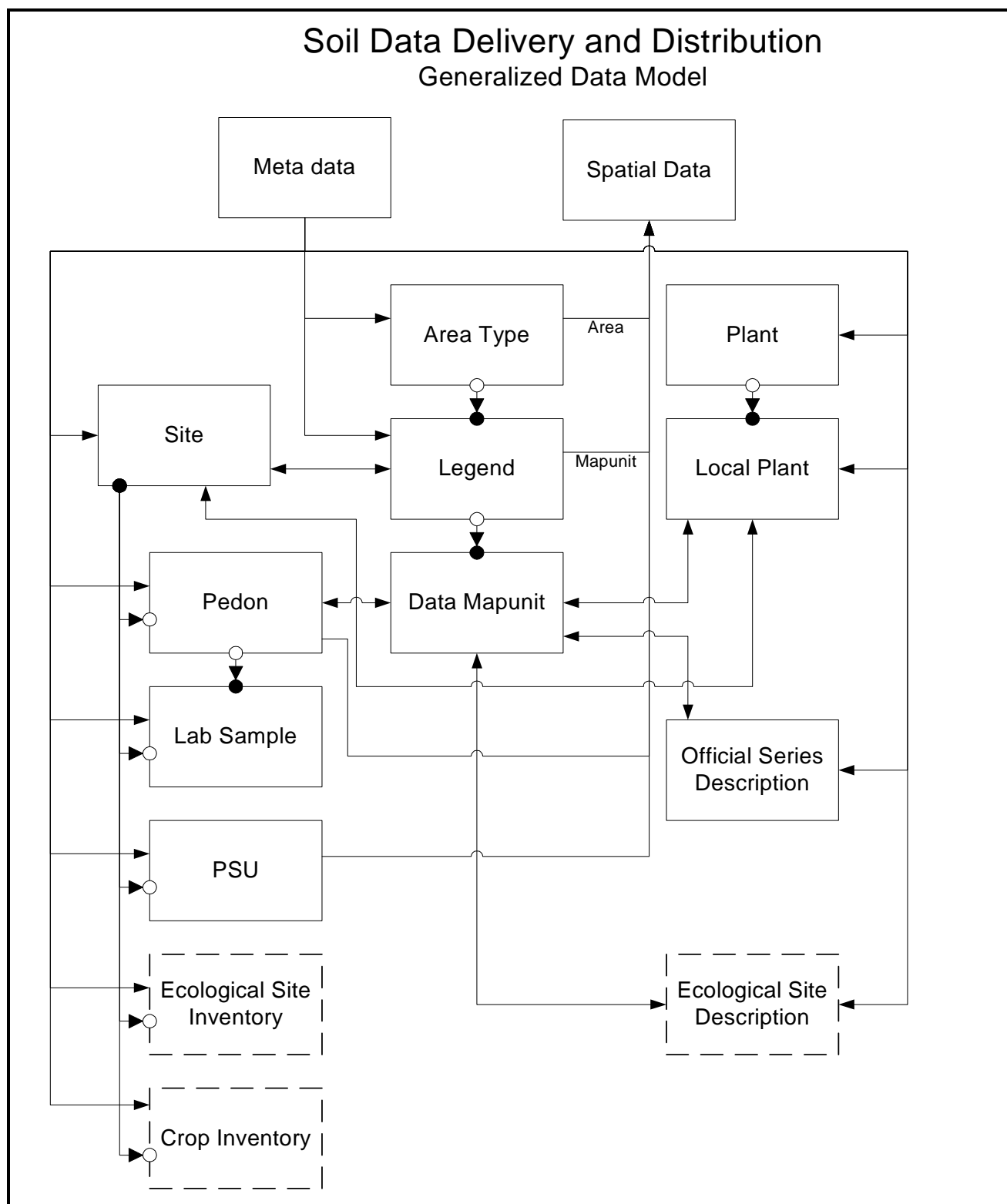
Review of Requirements

The following table lists the business requirements in priority order, as described in the DRS. Beneath each requirement are potential implementation options developed during a meeting held at the NCGC in Fort Worth November 14-15, 2001. These options guided the strategy described in this OPD, although not all of them are included in the recommended system. The column headed "Addressed by Design Unit" identifies which of the above listed OPD design units support the requirement.

Requirement with Implementation Options	Business Area	Addressed by Design Unit
6.20 - Provide stability in product content and delivery format	USFS, SSURGO, Models	
o Provide overlap in product – old and new – for a period of time, either a live export from NASIS in both formats or both formats in a warehouse. o Capability in warehouse to produce both formats. o Includes database formats and structure formats. o Provide data in one and only one data structure format. o Maintain an archive or snapshot of a dataset o Have software that can convert back to an older format.		o Warehouse Database o FOTG Data Mart
6.23 - Protect data from loss or modification	All	
o Provide adequate controls on official data, not editable. o Q&A review mechanisms to ensure what is delivered to the warehouse and customers is valid. o Have adequate backup and mirroring. o Ensure that any transformation of data to meet other requirements does not change data content. o Maintain spatial data resolution and accuracy as it is reformatted or projected.		o Warehouse Database o Staging Server
6.2 - Identify and access current official data for a specific use	FOTG, NRI, SSURGO	
o Default is most current as of this date, with options to specify a date or a particular use and get the "official" dataset for that date or use. o Capture old datasets for a particular use and store those in a warehouse for use (e.g., frozen soils data). o For certain uses of or vintages of soil data, provide a contact point (e.g., State Soil Scientist) to get access to that data.		o Warehouse Database o FOTG Data Mart o Web SDV o Reports Data Mart o Soils API
6.21 - Eliminate inconsistency	All	
o All products come from a single source of data. o If attributes do not change, but interpretive criteria changes causing different interpretations, do we have		o Warehouse Database

to store all versions of interps or replace old interp results with new ones or what? o Store interp results at the time of warehouse population. o Generate interp results using criteria in effect at the time of certification. o Create a yearbook of soil survey data (see Russ).		
6.11 - Deliver data to meet specific needs	SSURGO, FOTG, Models	
o Identify the specific needs. Datasets for these specific needs are already created (e.g., a data mart for a specific use). o Manage user interface to provide a list of options/needs for delivering data. o Limit the number of "needs" supported. o Use a tool that allows users to define their own export format/content. o Third party (vendor) provides this service.		o FOTG Data Mart o Custom Exports
6.22 - Eliminate redundancy	All	
o Minimize multiple formats for products that are currently manual processes difficult to keep consistent. o Applications access source data, not detached data. o Eliminate introduction of error in creating multiple formats and products through automation. o Use push technology to automatically update detached datasets.		o FOTG Data Mart o Soils API
6.3 - Identify and access previous versions of official data for a specific use	NRI, CRP	
o Keep versions of data in old and new formats along with interpretations and criteria. o Convert old data into new formats for a particular use.		o Warehouse Database o FOTG Data Mart o Reports Data Mart
6.17 - Provide complete national coverage of data	National Program	
o Use STATSGO o Use the soil survey area boundary layer tied to the area and legend tables in NASIS o Tabular data coverage for the entire country (?). o Manually determine which data is used and which data is not to avoid double counting.		o Warehouse Database
6.6 - Access to Reporting and downloading Capability	All	
o Use MUIR database o Pick from a list of canned reports.		o Reports Data Mart

o Provide custom report capability. o Select certain data elements and download those or format those into a report and print or download the report. o Download in ASCII delimited format. o Web Soil Data Viewer. o Query, download and print via a web interface. o API to shield actual warehouse from programmatic and user view. o Soil survey manuscript type reports pre-done or print on demand.		o Web SDV o Soils API o FOTG Data Mart o Custom Exports
6.12 - Generate new interpretations from current or previous official data	National Programs	
o NASIS Interpretation Generator or similar tool available within the warehouse. o Create new interps in NASIS and export criteria to warehouse. o Create new interps within warehouse and apply to warehouse data		o Reports Data Mart o Staging Server
6.13 - Apply interpretive criteria to selected map units or geographic areas	LESA, National Programs	
o NASIS Interpretation Generator or similar tool available within the warehouse. o Some capability to select map units – spatial or attribute based query		o Reports Data Mart
6.4 - Maintain, identify and access more than one set of official data for a geographic area	SSURGO STATSGO	
o Generate STATSGO from SSURGO. o Store multiple physical coverages. o Look at changes in the data model to allow a single base dataset with multiple interpretive results (*interp data object proposal*).		o Warehouse Database
6.16 - Provide data that can be used to create seamless spatial coverage	General	
o Process for matching survey areas spatially occurs prior to population of the data warehouse (eliminating slivers, overlaps, etc.). o Include mapping and correlation for a seamless join.		o Warehouse Database o Staging Server
6.26 - Provide Metadata, detailed information or description of products and data provided to users	All	
o Track version of every rule, evaluation, and property for interpretations, and version of the data dictionary. o Provide standard FGDC metadata file format. o Update all current metadata to latest FGDC standard automatically.		o Warehouse Database o FOTG Data Mart o Custom Export

o Include definitions of static metadata (choice lists) along with column definitions, etc. on the fly. o Generate different formats of metadata from single source. o Provide contact information for specific area of data (e.g., zoomed-in too far). o Provide interactive assistance for helping users get what they want.		
6.8 – Identify geographic areas where data have been updated since some previous date	NRI	
o Graphic display based on some user entered date. o Use archive date for survey area to display a list of surveys since a user entered date. o Click on a point and display date of last change.		o Warehouse Database o Reports Data Mart
6.24 - Comply with Federal Geographic Data Committee (FGDC) Soil Geographic Data Standard	All	
o Update current FGDC standard to match current database. o Provide at least one export format that meets FGDC standard.		o FOTG Data Mart
6.5 – Maintain, identify and provide access to soil survey supporting data for a geographic area	Tech Soil Services	
o Supporting data will be versioned. o Linkage information should change when a new version comes in. o Since NASIS key stays the same, need to add a version key to differentiate. o Provide information about which Access template to use with a dataset. o Ability to add information and data directly to the warehouse without going through NASIS (pictures, spatial, etc.).		o Warehouse Database o FOTG Data Mart o Staging Server
6.14 - Provide selected attributes for any geographic area	Tech Soil Services	
o Provide navigational interface to identify an area, and then provide an interface to choose appropriate attributes.		o Reports Data Mart o Custom Exports
6.7 – Identify changes between versions of data released to users	General	
o Provide a list at the object level of what changed without the details but include the date it changed. o Provide a detailed list of the changes attribute by attribute or polygon by polygon.		o Reports Data Mart
6.18 – Notify data users when data have changed	General	
o Allow users the ability to sign up for a notification if data of interest changes.		o FOTG Data Mart

|		
o Send notice to users who opt-in of changes to data. o Post a change list to a web site. o Track users when they access the site to notify them of changes of interest.		
6.10 – Deliver the most up-to-date data for specific uses	Tech Soil Services	
o This is in NASIS with all caveats.		o NASIS
6.27 - Integration with Other Resource Databases	All	
o Provide external linkages. o Provide an integrated database. o Geographic coincidence. o Provide some public search method for warehouse data. o Provide a means for populating the warehouse outside of NASIS (e.g., FS ??).		o Warehouse Database
6.25 – Provide access to the technical references, standards and guides for soil survey	All	
o Provide links to existing. o Integration of SC and OSD into NASIS as precursor to getting into a warehouse. o Maintain versions of these items??		o Warehouse Database
6.9 – Identify changes between the most up-to-date data and official data	State Soil Scientist	
o Provide compare capability between NASIS and warehouse. o Provide changes at the object level by date. o Provide detailed changes attribute by attribute.		o Potential addition to Reports Data Mart
6.19 – Notify data users of product plans and progress	General	
o Put up a web site with product plans and progress. o Enhanced status map with clickable links to state soil scientist. o State soil scientist provides status information for status map.		o Not directly addressed in OPD
6.15 – Select data by any attribute without respect to geographic area	Tech Sers, National Programs	
o Provide a query capability on any attribute (tabular or spatial) in the database, and graphically display results. o Define a geographic area by attribute.		o Potential addition to Reports Data Mart

Design Unit Sequence

One purpose for designating design units is to identify relatively independent groups of system capabilities that can be developed and deployed in a planned sequence of phases. In most large systems, resource or technical constraints preclude developing all the required capabilities at once. This section describes a proposed plan for phased development. Completion dates for each phase are highly uncertain, and depend on funding levels and priorities that are not yet determined.

Phase 1

Estimated completion: September, 2002

- Warehouse Database Design Unit.

- Interim procedure to export data from NASIS and import into warehouse database.
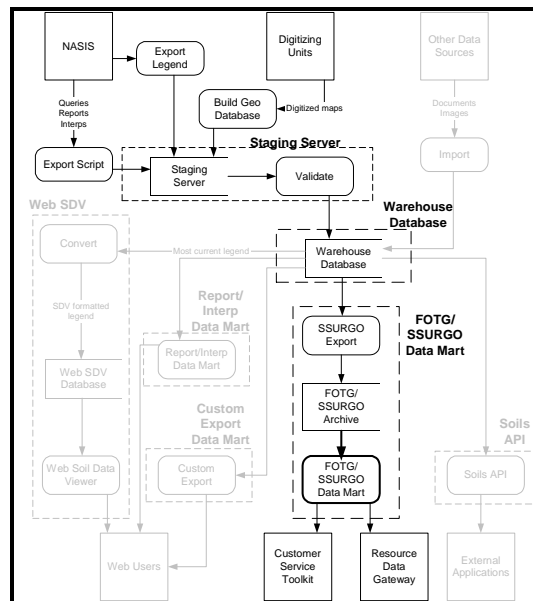
- FOTG / SSURGO Data Mart.
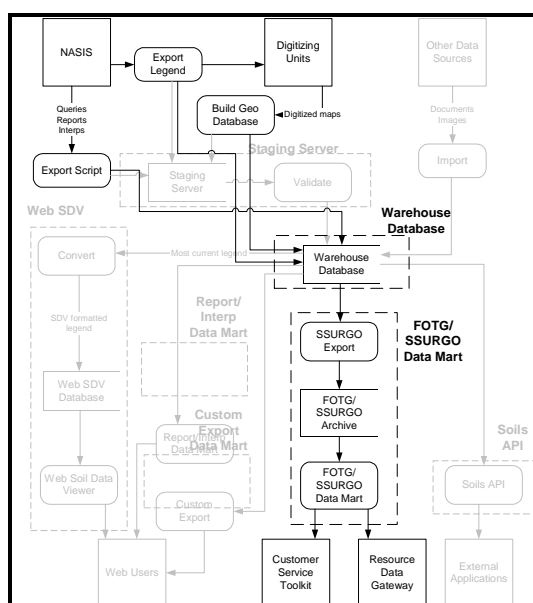
Cost Estimates

- Hardware/software for warehouse database: $200,000 to $300,000

- Hardware/software for FOTG Data Mart: $50,000 to $75,000

- Software development and conversion: $250,000 to $400,000

- Total: $500,000 to $775,000

Phase 2

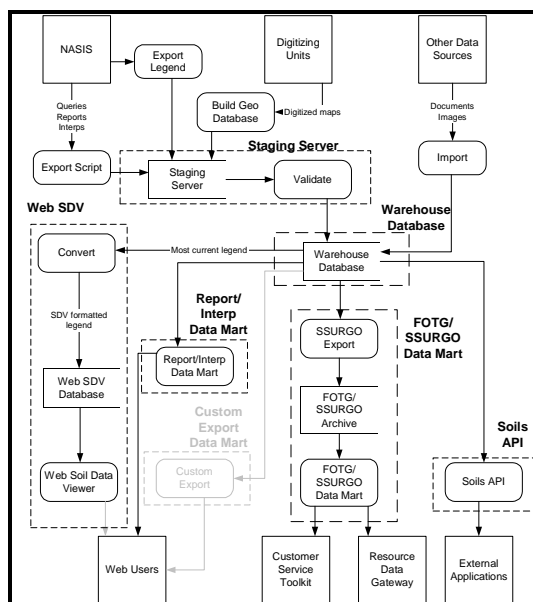Estimated completion: September, 2002

- Staging Server Design Unit.

Cost Estimates

- Hardware/software for staging server: $50,000 to $75,000

- Software development and conversion: $50,000 to $100,000

- Total: $100,000 to $175,000

Phase 3

Estimated completion: March, 2003

- Reports and Interpretations Data Mart, initial implementation including generation of interpretations and reports using NASIS style scripts.

- Web Soil Data Viewer.

- Soils API, initial implementation using ODBC access.
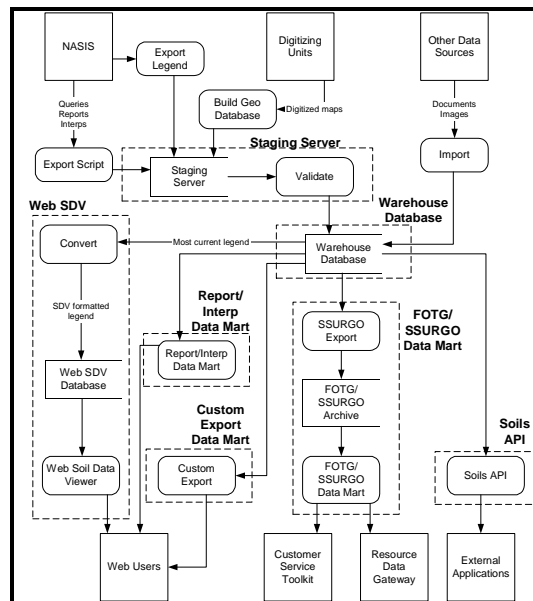

Cost Estimates

- Hardware/software for Web SDV server: $100,000 to $175,000

- Hardware/software for Reports and API server: $50,000 to $75,000

- Software development and conversion: $100,000 to $150,000

- Total: $250,000 to $400,000

Phase 4

Estimated completion: September, 2003

- Custom Data Export Data Mart

- Version comparison report

- Soils API with enhanced business rule support

Cost Estimates

- Software development and conversion: $100,000 to $150,000
- Total: $100,000 to $150,000

Conversion Approach

The new data delivery system will replace certain functions performed by existing systems, so a transition plan needs to be developed. This section describes a general approach to the transition issues, and more detailed plans will be developed in a later stage of the analysis.

SSURGO Certification

The Soil Survey Geographic (SSURGO) Database consists of spatial data, tabular data, and metadata, which are currently collected and verified by the Digitizing Units in a procedure referred to as SSURGO certification. To meet the needs of a wide variety of users, the data are distributed in several formats.

- The spatial data are distributed in four formats: Digital Line Graph Optional (DLG-3) format, ArcInfo coverages, ArcInfo Interchange (.e00) files, and ArcView shapefiles. The spatial data are also distributed in 3.75 or 7.5 minute quadrangle format and as a seamless survey area.

- The tabular data, exported from NASIS, are distributed in three formats: in delimited text format, as INFO tables, and as ArcInfo Interchange (.e00) files (By a recent policy decision, support for the INFO table format will be dropped).

- The metadata file is distributed in two formats: in text format and as HTML.

Distributing spatial data in this many formats does not comply with the objectives of a soil data delivery system. It does not provide stability in delivery content, eliminate inconsistency, or most notably, eliminate redundancy. We recommend that the redesign of the SSURGO continue to address the question of which formats need to be provided.

With a proposal of a true "geospatial database" containing both spatial and tabular data, the stated objectives could be met. The data would only need to be stored in one format with the warehouse being capable of producing other desired formats. The most important step in the SSURGO certification would be creating the link or validating the referential integrity between the spatial and tabular data (the SSURGO Certification AMLs). The procedures now used to produce the output formats (SSURGO Distribution AMLs) would be changed to use the warehouse as a data source. The products would be distributed through the FOTG / SSURGO Data Mart.

Existing SSURGO 2.0 Datasets

SSURGO products are now being produced in the new Version 2 format, and that format is not expected to change during the transition to the new delivery system. It would be desirable to avoid the workload of re-certifying those legends that are already certified; in some cases the data in NASIS have changed and it would not be possible to exactly recreate the SSURGO export.

Three general approaches to this problem have been identified: 1) keep current SSURGO files in the FOTG Data Mart and not in the warehouse, 2) freeze the NASIS data for existing exports until they can be copied to the warehouse, or 3) import current SSURGO files into the

warehouse. These approaches will be examined further in the next stage of analysis, and management decisions will be requested before implementation.

1. If existing SSURGO data sets are incorporated into the FOTG Data Mart they will be available for download even if the same surveys are not in the warehouse. Some data distribution functions, such as reporting and interpretation, would not be available for these surveys until a new data set is exported from NASIS.

2. After a SSURGO export from NASIS the soil survey data objects in NASIS could be "frozen" until the data are exported to the warehouse. This could be implemented by changing the ownership of the data after an export. Interpretation criteria could still change during the period from the SSURGO export to warehouse population, but this could be addressed by copying interpretation results from the SSURGO export back into the warehouse.

3. A more costly solution would be to develop a procedure to import existing certified SSURGO data sets into the warehouse database. The design of such a procedure is complicated by the fact that a SSURGO export may exclude some NASIS data, such as minor components and additional mapunit symbols. Warehouse entries derived from SSURGO would not be the same as those from NASIS. A management decision would be needed as to whether this is acceptable, given that the data delivery goals of the warehouse are similar to those of SSURGO.

    An additional complication is that the data model for the warehouse will be more like NASIS than SSURGO. To import from SSURGO to the warehouse some missing linkages, such as the Correlation table, have to be created. This could be done through an automated process to match a SSURGO export to the data in NASIS and find the additional data needed to fully populate the warehouse. Since recent SSURGO exports are not likely to have changed much in NASIS, this should be feasible in most cases, but manual intervention will be needed in some cases. At best an import from SSURGO would only approximately match what would come from NASIS.

Seamless Coverage

A DRS requirement is to support the creation of a nation wide soil map layer without ambiguities at the borders of survey areas. Existing digitized soil maps do not generally meet this requirement, having gaps and overlaps in adjoining polygons from independently digitized maps. There are also inconsistencies in the tabular data across survey area boundaries. Automated means of correcting for these problems are not considered acceptable, because decisions by a soil scientist familiar with the area in question are needed.

The SDE product allows retention of gaps and overlaps among the polygons, so there is no technical requirement to correct these problems before creating the warehouse database. The recommended course of action is to place existing digitized maps into the warehouse as is, and develop a plan to resolve edge matching issues over time.

Summary of Conclusions

This OPD describes a strategy for developing a soil data delivery and distribution system, including a phased implementation plan and accompanying decision points. The following list summarizes in priority order the key features of this strategy and identifies some business decisions that will be required during the next phases of the project.

- A warehouse database will be constructed using Informix and SDE to serve as a single source of official soil survey data.

- Interpretations selected by the State Soil Scientist will be stored with the soil survey data in the warehouse.

- The SSURGO certification process will change to support the population of the warehouse with digitized maps and NASIS tabular data. Preparation of products for delivery will be a separate process in a data mart.

- SSURGO data sets that are certified before the warehouse begins operation will be available through the same data mart. There are issues to be resolved about how these data will appear in the warehouse.

- The system will be capable of storing and delivering most of the data for the Field Office Technical Guide, Section 2, if the states choose to use it that way. Some aspects of the FOTG support will require further consideration.

- Additional interpretations and reports can be generated from data in the warehouse and delivered to users.

- The system will store old versions of soil data, generate reports and interpretations from old versions, and report differences between versions for a given location. The method of managing versions will be designed on the assumption that a request for a specific version of data is more time critical than a request for the differences between versions.

- The system will provide at least one data export option that is FGDC compliant.

- Online viewing and analysis of soil maps will be available to the public through the Web Soil Data Viewer.

- Soil characterization data will be available from the system, but the process for obtaining these data from soils labs is not yet designed.

- Soil classification and official series descriptions will also be available from the system, and further work is needed to determine how this will be integrated into the database.

- Models and other external applications will be able to access the system to obtain needed data. There are several ways this could be accomplished, at different levels of sophistication and development cost.